Homework 6

March 29, 2022

Jose Carlos Munoz

1

The RNN would most likely suffer from vanishin gradient. While the transformation of the function is 3.5, the derivative of the sigmoid activation is always less than 0.25. When doing back propogation, the value between these numbers is less than 1. So , in a Recurrent network, it will be dramatically go small. So this RNN will suffer from vanishing gadient.

2

The rule is that spectral radius times the activation function highest value to be as close to one. Since, we know that the value can blow up 8 times high, we can set the spectral radius to a $\frac{1}{8}$. So that they both multiply to be at most 1

3

The computation graph of the RNN cell is as follows